Optimization on Quadratic Matrix Lie Groups

Luís Miguel Machado*
Departamento de Matemática
Universidade de Trás-os-Montes e Alto Douro
5001-911 Vila Real - Portugal
lmiguel@utad.pt

F. Silva Leite *
Departamento de Matemática
Universidade de Coimbra
3001-454 Coimbra - Portugal
fleite@mat.uc.pt

Abstract

We study optimality properties of the smooth function $\mathrm{tr}\left(\Theta^{-1}Q\Theta N - 2M\Theta^{-1}\right)$, viewed as a function of Θ, with Θ belonging to certain quadratic matrix Lie groups which are generalizations of the orthogonal group. Some optimization matrix problems are formulated in terms of this function. Computational issues based on continuous algorithms are discussed.

Keywords: Lie group, Lie algebra, Riemannian metric, gradient vector field, gradient flow, complete integrability.

Introduction

The theory of completely integrable Hamiltonian systems has had a great development in recent years. In [1], Bloch showed that some of those systems are closely related to gradient flows of smooth functions. In this paper we study optimality properties of the smooth function

$$\mathrm{tr}\left(\Theta^{-1}Q\Theta N - 2M\Theta^{-1}\right),$$

viewed as a function of Θ, with Θ belonging to the Lie group of $P-$orthogonal matrices G. For a convenient choice of the matrix P, Cardoso and Silva Leite ([4]) proved that the Lie group of $P-$orthogonal matrices is isomorphic to one of the classical Lie groups $O(p,q)$, where $p,\ q \in \mathbb{N}_0$ and $p+q=n$, or $Sp(n,\mathbb{R})$ (the symplectic group).

It is well known (see [7]) that, for an appropriate choice of a Riemannian metric on G, the gradient flow equation of the smooth function $\eta : G \longrightarrow \mathbb{R}$ is the system of ordinary nonlinear

*Work supported in part by ISR and research network contract FMRXCT-970137.

differential equations

$$\dot{\Theta} = \nabla \eta \left(\Theta \right). \tag{1}$$

The gradient flow equation can be viewed, not only, as a great geometric mathematical tool, but also, as a simulation tool that can be used in engineering to solve some standard problems on applied mathematics. In fact, there are explicit algebraic methods available to solve various computational problems that are relatively easy to implement on a digital computer. However, we present here the continuous time varying system (1), whose limiting solution solve the specific computational problem. Working with various discretizations of the continuous time system (1), it opens the possibility of finding new and perhaps faster algorithms to solve a given problem than a purely algebraic approach would allow.

Actually, the gradient flow equation (1) evolves on the Lie group G and provides a stable algorithm that converges to an equilibrium point of (1).

For the particular case when $P = I$, Brockett [2, 3] already proved that the gradient flow can be used to solve certain computational problems, like diagonalize symmetric matrices, sort lists of real numbers and solve some linear programming problems. Bloch [1] also showed that the Toda lattice problem and the total least squares problem could be recast as gradient flows.

An interesting result given here makes contact with the work of Munthe-Kaas et al [9] and provides a way of finding, for $P^2 = \pm I$, the $P-$orthogonal matrix that is nearest to a given nonsingular matrix, by using a gradient flow equation.

1 Properties of $P-$orthogonal Matrices

In what follows $\mathcal{G}l(n, \mathbb{R})$ denote the Lie algebra of all n by n matrices with real entries and $GL(n, \mathbb{R})$ the Lie group of all n by n nonsingular matrices.

Let P be any n by n orthogonal matrix ($P^\top = P^{-1}$) and

$$G = \{ \Theta \in GL(n, \mathbb{R}) \, : \, \Theta^\top P \Theta = P \} . \tag{2}$$

G is an algebraic closed subgroup of $GL(n, \mathbb{R})$ and so, is itself a Lie group.

Remark 1 *Since G depends on P, it would be more appropriate to denote it by G_P. However, for the sake of simplicity we stick to that notation.*

Lemma 1 *$\Theta \in G$ iff $\Theta^\top \in G$.*

Proof:

$$\Theta \in G \quad \Longleftrightarrow \quad \Theta^\top P \Theta = P \iff \Theta^\top = P \Theta^{-1} P^\top$$

$$
\begin{aligned}
&\iff \quad \Theta = P\left(\Theta^{-1}\right)^{\top} P^{\top} \iff \Theta P \Theta^{\top} = P \\
&\iff \quad \left(\Theta^{\top}\right)^{\top} P \Theta^{\top} = P \iff \Theta^{\top} \in G.
\end{aligned}
$$

■

The Lie algebra of G is the set

$$
\mathcal{L} = \left\{\Omega \in \mathcal{G}l(n, \mathbb{R}) \;:\; \Omega^{\top} P = -P\Omega\right\}, \tag{3}
$$

where the Lie bracket is the usual commutator of matrices ($[A,B] = AB - BA,\ A,\ B \in \mathcal{G}l(n,\mathbb{R})$).

Lemma 2 *$\Omega \in \mathcal{L}$ iff $\Omega^{\top} \in \mathcal{L}$.*

Proof:

$$
\begin{aligned}
\Omega \in \mathcal{L} \quad &\iff \quad \Omega^{\top} P = -P\Omega \iff \Omega^{\top} = -P\Omega P^{\top} \\
&\iff \quad \Omega = -P\Omega^{\top} P^{\top} \iff \Omega P = -P\Omega^{\top} \\
&\iff \quad \left(\Omega^{\top}\right)^{\top} P = -P\Omega^{\top} \iff \Omega^{\top} \in \mathcal{L}.
\end{aligned}
$$

■

Let us now characterize the tangent space of G at Θ.

Since G is a Lie group, the left translation

$$
\begin{array}{rccc}
L_{\Theta}\colon & G & \longrightarrow & G \\
 & \Delta & \longmapsto & \Theta\Delta
\end{array},
$$

defines a diffeomorphism on G and then, the tangent map of L_{Θ} at the identity matrix, I, defines the linear isomorphism between tangent spaces

$$
\begin{array}{rccc}
T_I L_{\Theta}\colon & T_I G & \longrightarrow & T_{\Theta} G \\
 & \Omega & \longmapsto & \Theta\Omega
\end{array},
$$

and thus

$$
T_{\Theta} G = \left\{\Theta\Omega \;:\; \Omega \in \mathcal{L}\right\}. \tag{4}
$$

We could also define the tangent space to G at Θ as being the set

$$
T_{\Theta} G = \left\{\Omega\Theta \;:\; \Omega \in \mathcal{L}\right\}, \tag{5}
$$

if we had considered the right translation on G, $R_{\Theta} : \Delta \in G \longmapsto \Delta\Theta$.

If we consider the set of matrices

$$
\mathcal{J} = \left\{A \in \mathcal{G}l(n, \mathbb{R}) \;:\; A^{\top} P = PA\right\}, \tag{6}
$$

it is easy to see that $\mathcal{J}$ is not closed under the commutator and so does not have the structure of a Lie algebra.

However, it can be shown that $\mathcal{J}$ has the structure of a Jordan algebra under the product $\{A, B\} = AB + BA$ ([10]).

Lemma 3 *$A \in \mathcal{J}$ iff $A^\top \in \mathcal{J}$.*

Proof:

$$\begin{aligned} A \in \mathcal{J} \quad &\Longleftrightarrow \quad A^\top P = PA \iff A^\top = PAP^\top \\ &\Longleftrightarrow \quad A = PA^\top P^\top \iff AP = PA^\top \\ &\Longleftrightarrow \quad (A^\top)^\top P = PA^\top \iff A^\top \in \mathcal{J}. \end{aligned}$$

$\blacksquare$

It can also be proved that

$$\{\mathcal{J}, \mathcal{J}\} \subset \mathcal{J}, \quad \{\mathcal{J}, \mathcal{L}\} \subset \mathcal{L} \text{ and } \{\mathcal{L}, \mathcal{L}\} \subset \mathcal{J}, \tag{7}$$

and that

$$[\mathcal{L}, \mathcal{L}] \subset \mathcal{L}, \quad [\mathcal{L}, \mathcal{J}] \subset \mathcal{J} \text{ and } [\mathcal{J}, \mathcal{J}] \subset \mathcal{L}. \tag{8}$$

We will refer to the matrices of G as $P-$orthogonal, the matrices of $\mathcal{L}$ as $P-$skew-symmetric and matrices of $\mathcal{J}$ as $P-$symmetric.

Note that if $P = I$, G coincides with the Lie group of orthogonal matrices $(O(n))$, $\mathcal{L}$ with the Lie algebra of skew-symmetric matrices $(so(n))$ and $\mathcal{J}$ with the Jordan algebra of symmetric matrices.

Let us define the $P-$transpose of a matrix $X \in \mathcal{G}l(n, \mathbb{R})$ as

$$X^P = P^\top X^\top P. \tag{9}$$

With this terminology we say that X is $P-$orthogonal if $X^P X = I$, X is $P-$skew-symmetric if $X^P = -X$ and X is $P-$symmetric if $X^P = X$.

We now present some elementary properties about the $P-$transpose of matrices.

Lemma 4 *If $X, Y \in \mathcal{G}l(n, \mathbb{R})$ then*

(i) $(X + Y)^P = X^P + Y^P$;

(ii) $(XY)^P = Y^P X^P$;

(iii) $[X, Y]^P = [Y^P, X^P]$;

(iv) $\{X,Y\}^P = \{X^P, Y^P\}$.

Proof:

(i) $(X+Y)^P = P^\top (X+Y)^\top P = P^\top X^\top P + P^\top Y^\top P = X^P + Y^P$;

(ii) $(XY)^P = P^\top Y^\top X^\top P = P^\top Y^\top P P^\top X^\top P = Y^P X^P$;

(iii) $[X,Y]^P = (XY - YX)^P = (XY)^P - (YX)^P$.

The result is now an immediate consequence of (ii).

(iv) Since $\{X,Y\}^P = (XY+YX)^P$, the result follows from (i) and (ii).

■

Lemma 5 $X \in \mathcal{J}\ (\mathcal{L}) \iff X^P \in \mathcal{J}\ (\mathcal{L})$.

Proof: Let us suppose that $X \in \mathcal{J}$. The case $X \in \mathcal{L}$ is treated in a similar way.

$$\begin{aligned} X^P \in \mathcal{J} \quad & \iff \quad (X^P)^\top P = PX^P \iff (P^\top X^\top P)^\top P = PP^\top X^\top P \\ & \iff \quad P^\top XPP = X^\top P \iff P^\top XP = X^\top \\ & \iff \quad XP = PX^\top \iff X^\top \in \mathcal{J} \\ & \iff \quad X \in \mathcal{J} \quad \text{(lemma 3)}. \end{aligned}$$

■

Lemma 6 *If* $\Theta \in G$ *and* $X \in \mathcal{G}l(n, \mathbb{R})$ *then,* $\left(\Theta^{-1} X \Theta\right)^P = \Theta^{-1} X^P \Theta$.

Proof: According to (9), and since $\Theta^\top P \Theta = P$, we have

$$\left(\Theta^{-1} X \Theta\right)^P = P^\top \Theta^\top X^\top \left(\Theta^{-1}\right)^\top P = \Theta^{-1} P^\top X^\top P \Theta = \Theta^{-1} X^P \Theta.$$

■

Lemma 7 *If* $\Theta \in G$ *and* $X \in \mathcal{J}\ (\mathcal{L})$ *then* $\Theta^{-1} X \Theta \in \mathcal{J}\ (\mathcal{L})$.

Proof: Let us consider $\Theta \in G$ and $X \in \mathcal{J}$. The case $X \in \mathcal{L}$ is analogous.

Then,

$$\left(\Theta^{-1} X \Theta\right)^\top P = \Theta^\top X^\top \left(\Theta^{-1}\right)^\top P = \Theta^\top X^\top P \Theta = \Theta^\top P X \Theta = P \Theta^{-1} X \Theta.$$

■

2 A Riemannian metric on G

In this section we assume that P is an orthogonal n by n matrix such that $P^2 = \pm I$.

Let us consider on the Lie group $GL(n, \mathbb{R})$, the automorphism σ defined by

$$\begin{array}{rccc} \sigma\colon & GL(n,\mathbb{R}) & \longrightarrow & GL(n,\mathbb{R}) \\ & \Theta & \longmapsto & \sigma(\Theta) = P^\top \left(\Theta^{-1}\right)^\top P = \left(\Theta^{-1}\right)^P \end{array}. \tag{10}$$

Since,

$$\begin{aligned} \sigma(\sigma(\Theta)) &= \sigma\left(P^\top \left(\Theta^{-1}\right)^\top P\right) = P^\top \left[\left(P^\top \left(\Theta^{-1}\right)^\top P\right)^{-1}\right]^\top P \\ &= P^\top \left(P^\top \Theta^\top P\right)^\top P = P^\top P^\top \Theta P P \\ &= \Theta, \end{aligned}$$

σ is an involutive automorphism on $GL(n, \mathbb{R})$.

Let us consider the subsets of $GL(n, \mathbb{R})$ defined by

$$G^\sigma = \left\{\Theta \in GL(n,\mathbb{R}) \,:\, \sigma(\Theta) = \Theta^{-1}\right\}, \tag{11}$$

and

$$G_\sigma = \left\{\Theta \in GL(n,\mathbb{R}) \,:\, \sigma(\Theta) = \Theta\right\}. \tag{12}$$

Lemma 8 *$G^\sigma = \mathcal{J} \cap GL(n, \mathbb{R})$ and $G_\sigma = G$.*

Proof: Attending to the definition of σ,

$$\sigma(\Theta) = \Theta^{-1} \iff P^\top \left(\Theta^{-1}\right)^\top P = \Theta^{-1} \iff P^\top \Theta^\top P = \Theta \iff \Theta^\top P = P\Theta \iff \Theta \in GL(n,\mathbb{R}) \cap \mathcal{J},$$

and

$$\sigma(\Theta) = \Theta \iff P^\top \left(\Theta^{-1}\right)^\top P = \Theta \iff P^\top \Theta^\top P = \Theta^{-1} \iff \Theta^\top P \Theta = P \iff \Theta \in G.$$

■

Although the set G^σ does not have a group structure, it can be shown that it is a symmetric space (see Munthe-Kaas M. et al [9]).

We can also define similar sets on the Lie algebra of $GL(n, \mathbb{R})$ if we consider the tangent map of σ

$$\begin{array}{rccc} T_I\sigma\colon & \mathcal{G}l(n,\mathbb{R}) & \longrightarrow & \mathcal{G}l(n,\mathbb{R}) \\ & X & \longmapsto & T_I\sigma(X) = \left.\dfrac{d}{dt}\right|_{t=0} \sigma\left(\mathrm{e}^{tX}\right) = -P^\top X^\top P = -X^P \end{array}. \tag{13}$$

Since σ is an involutive automorphism on $GL(n, \mathbb{R})$, $T_I\sigma$ is also an involutive automorphism on $\mathcal{G}l(n, \mathbb{R})$. If λ is an eigenvalue of $T_I\sigma$ associated to the nonzero eigenvector X, then,

$$X = T_I\sigma\left(T_I\sigma(X)\right) = T_I\sigma(\lambda X) = \lambda^2 X,$$

and thus

$$\lambda = \pm 1.$$

Let us consider the vector subspaces of $\mathcal{G}l(n, \mathbb{R})$

$$\begin{aligned} \mathcal{T}_\sigma &= \{X \in \mathcal{G}l(n, \mathbb{R}) \,:\, T_I\sigma(X) = X\}\,, \\ \mathcal{P}_\sigma &= \{X \in \mathcal{G}l(n, \mathbb{R}) \,:\, T_I\sigma(X) = -X\}\,, \end{aligned} \tag{14}$$

corresponding to the eigenvalues 1 and -1, respectively.

Lemma 9 *$\mathcal{T}_\sigma = \mathcal{L}$ and $\mathcal{P}_\sigma = \mathcal{J}$.*

Proof: Follows immediately from the definition of the tangent map $T_I\sigma$. ■

Associated to the involutive automorphism $T_I\sigma$ we have the direct sum decomposition of $\mathcal{G}l(n, \mathbb{R})$ (see [9])

$$\mathcal{G}l(n, \mathbb{R}) = \mathcal{L} \oplus \mathcal{J},$$

and thus, we can write any element X of $\mathcal{G}l(n, \mathbb{R})$ as

$$X = \frac{X - X^P}{2} + \frac{X + X^P}{2},$$

where $X - X^P \in \mathcal{L}$ and $X + X^P \in \mathcal{J}$. Moreover, this decomposition is unique and is nothing else than the canonical decomposition of a matrix into its $P-$skew-symmetric and $P-$symmetric part (Silva Leite F. and Peter E. Crouch [10]).

The involutive automorphism on $\mathcal{G}l(n, \mathbb{R})$, (13), enables us to define a canonical inner product on a semisimple Lie subalgebra of $\mathcal{G}l(n, \mathbb{R})$.

The Killing form, also known as the Cartan inner product, is the symmetric bilinear form defined by

$$\begin{array}{llll} K\colon & \mathcal{G}l(n, \mathbb{R}) \times \mathcal{G}l(n, \mathbb{R}) & \longrightarrow & \mathbb{R} \\ & (X, Y) & \longmapsto & K(X, Y) = \mathrm{tr}(ad_X\, ad_Y), \end{array}$$

where ad_X represents the adjoint action on $\mathcal{G}l(n, \mathbb{R})$ defined by $ad_X(Y) = [X, Y]$.

Given a linear subspace $\mathcal{G}$ of $\mathcal{G}l(n, \mathbb{R})$, the orthogonal complement of $\mathcal{G}$ relatively to K is the subspace

$$\mathcal{G}^\perp = \{Y \in \mathcal{G}l(n, \mathbb{R}) \,:\, K(X, Y) = 0,\, \forall X \in \mathcal{G}\}\,.$$

If $\mathcal{G}$ is a semisimple Lie subalgebra of $\mathcal{G}l(n, \mathbb{R})$ then K is nondegenerate on $\mathcal{G}$, which means that, if $X \in \mathcal{G}$ is such that $K(X, Y) = 0$, $\forall Y \in \mathcal{G}$, then $X = 0$.

Theorem 1 *Let $\mathcal{G}$ be a semisimple Lie subalgebra of $\mathcal{G}l(n, \mathbb{R})$ invariant under the involutive automorphism (13) and consider the Cartan decomposition*

$$\mathcal{G} = (\mathcal{L} \cap \mathcal{G}) \oplus (\mathcal{J} \cap \mathcal{G})\,. \tag{15}$$

Then the bilinear form

$$\langle X,Y\rangle_P = K\left(X, P^\top Y^\top P\right) = K\left(X, Y^P\right), \;\; X,\, Y \in \mathcal{G}, \tag{16}$$

is an inner product on $\mathcal{G}$ and the subspaces $\mathcal{L}\cap\mathcal{G}$ and $\mathcal{J}\cap\mathcal{G}$ are orthogonal with respect to this inner product.

Proof: Since $\mathcal{G}$ is semisimple, the Killing form is nondegenerate in $\mathcal{G}$. Furthermore, automorphisms of $\mathcal{G}$ leave the Killing form invariant so, since (13) is involutive, the bilinear form (16) is symmetric. According to (15), given an arbitrary $Z \in \mathcal{G}$, it can be uniquely written as

$$Z = X + Y,$$

with $X \in \mathcal{L}\cap\mathcal{G}$ and $Y \in \mathcal{J}\cap\mathcal{G}$.

Thus,

$$\begin{aligned}
\langle Z,Z\rangle_P &= K\left(X+Y, (X+Y)^P\right) \\
&= K\left(X, X^P\right) + K(X,Y^P) + K(Y,X^P) + K(Y,Y^P) \\
&= -K\left(X,X\right) + K(X,Y) - K(Y,X) + K(Y,Y) \\
&= -K(X,X) + K(Y,Y).
\end{aligned}$$

But (15) is a Cartan decomposition of $\mathcal{G}$, thus the Killing form K is negative semidefinite on $\mathcal{L}\cap\mathcal{G}$ and positive semidefinite on $\mathcal{J}\cap\mathcal{G}$, which means that $\langle Z,Z\rangle_P \geq 0$ and is equal to zero if and only if $Z=0$.

Let us now prove that the subspaces $\mathcal{L}\cap\mathcal{G}$ and $\mathcal{J}\cap\mathcal{G}$ are orthogonal with respect to (16).

If $X \in \mathcal{L}\cap\mathcal{G}$ and $Y \in \mathcal{J}\cap\mathcal{G}$ then,

$$\begin{aligned}
\langle X,Y\rangle_P &= K\left(X, P^\top Y^\top P\right) = K\left(P^\top X^\top P, Y\right) \\
&= -K\left(P^\top P X, Y\right) = -K\left(X, P^\top P Y\right) \\
&= -K\left(X, P^\top Y^\top P\right) = -\langle X,Y\rangle_P.
\end{aligned}$$

■

By computing explicitly the Killing form on the semisimple Lie algebra $\mathcal{G}$ (see Helgason [6]), the inner product (16) is given, to within a scale factor, by

$$\langle X,Y\rangle_P = \operatorname{tr}\left(XP^\top Y^\top P\right) = \operatorname{tr}\left(X^P Y\right), \;\; X,\, Y \in \mathcal{G}. \tag{17}$$

From now on we will denote by $\mathcal{L}$ and $\mathcal{J}$ the sets $\mathcal{L}\cap\mathcal{G}$ and $\mathcal{J}\cap\mathcal{G}$, respectively. We keep the terminology G to denote a Lie group of $P-$orthogonal matrices having $\mathcal{L}$ as Lie algebra.

Remark 2 *We note that when $P = I$, $\langle \cdot, \cdot \rangle_P$ is the Frobenius inner product.*

Since we can endowed any subspace of $\mathcal{G}$ with the inner product (17), the restriction of (17) to the Lie algebra $\mathcal{L}$ carries out a Riemannian structure to the Lie group G, if we define in the tangent space of G at Θ, $T_\Theta G$, the inner product

$$\ll \Theta\Omega_1, \Theta\Omega_2 \gg = \langle \Omega_1, \Omega_2 \rangle_P, \;\; \Omega_1, \; \Omega_2 \in \mathcal{L}. \tag{18}$$

Remark 3 *If we consider the extension of the inner product (17) to the Lie algebra $\mathcal{G}l(n, \mathrm{I\!R})$, the positive definiteness condition gets lost. However, it defines a pseudo-Riemannian metric on the manifold of nonsingular matrices $GL(n, \mathrm{I\!R})$. The restriction of that bilinear form to the Lie algebra $\mathcal{L}$ of a closed Lie subgroup of $GL(n, \mathrm{I\!R})$ gives a pseudo-Riemannian structure to that subgroup.*

Remark 4 *The results presented in this section were obtained under the assumption that P must be an orthogonal matrix satisfying $P^2 = \pm I$. This last condition may seems to be a significant restriction. However, in a recent work, Cardoso J. R. and F. Silva Leite [4] proved that if P is either symmetric, or skew-symmetric the Lie group G is isomorphic to one of them. More precisely, if P is a symmetric matrix having p positive eigenvalues and q negative eigenvalues, the Lie group G is isomorphic to the orthogonal Lie group*

$$O(p, q) = \left\{ \Theta \in GL(n, \mathrm{I\!R}) : \Theta^\top D \Theta = D \right\},$$

where

$$D = \begin{bmatrix} I_p & 0 \\ 0 & -I_q \end{bmatrix},$$

with p, $q \in \mathrm{I\!N}_0$ and $p + q = n$, and if P is skew-symmetric, with $n = 2m$ even, G is isomorphic to the symplectic group

$$Sp(n, \mathrm{I\!R}) = \left\{ \Theta \in GL(n, \mathrm{I\!R}) : \Theta^\top J \Theta = J \right\},$$

where

$$J = \begin{bmatrix} 0 & I_m \\ -I_m & 0 \end{bmatrix},$$

and I_d stands for the d by d identity matrix.

3 The Generalized Brockett Function

Let us consider the function η, defined on the Lie group of $P-$orthogonal matrices G ($P^\top = P^{-1}$),

$$\begin{array}{rccl} \eta\colon & G & \longrightarrow & \mathrm{I\!R} \\ & \Theta & \longmapsto & \eta(\Theta) = \operatorname{tr}\left(\Theta^{-1} Q \Theta N - 2M\Theta^{-1}\right) \end{array}, \tag{19}$$

where Q, M and N are given n by n matrices and tr denotes, as usual, the trace function.

We will refer to this function as the generalized Brockett function, since it is the generalization of the Brockett function ([3]) for the Lie group of $P-$orthogonal matrices.

In order to find the stationary points of η, let us compute the tangent map of η at Θ.

Since

$$\begin{array}{rccl} \gamma\colon & I \subset \mathbb{R} & \longrightarrow & G \\ & t & \longmapsto & \gamma(t) = \Theta \mathrm{e}^{t\Omega} \end{array},$$

defines a smooth curve on G that passes through Θ at $t = 0$ and whose velocity vector at the same instant is given by $\Theta\Omega \in T_\Theta G$, the tangent map of η at Θ can be defined as

$$\begin{array}{rccl} T_\Theta \eta\colon & T_\Theta G & \longrightarrow & \mathbb{R} \\ & \Theta\Omega & \longmapsto & T_\Theta \eta(\Theta\Omega) = \left.\dfrac{d}{dt}\right|_{t=0} (\eta \circ \gamma)(t) \end{array}.$$

Then,

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=0} (\eta \circ \gamma)(t) &= \left.\frac{d}{dt}\right|_{t=0} \operatorname{tr}\left(\left(\Theta \mathrm{e}^{t\Omega}\right)^{-1} Q \Theta \mathrm{e}^{t\Omega} N - 2M \left(\Theta \mathrm{e}^{t\Omega}\right)^{-1}\right) \\ &= \left.\frac{d}{dt}\right|_{t=0} \operatorname{tr}\left(\mathrm{e}^{-t\Omega}\Theta^{-1} Q \Theta \mathrm{e}^{t\Omega} N - 2M \mathrm{e}^{-t\Omega}\Theta^{-1}\right) \\ &= \operatorname{tr}\left(-\Omega\Theta^{-1} Q \Theta N + \Theta^{-1} Q \Theta \Omega N + 2M\Omega\Theta^{-1}\right) \\ &= \operatorname{tr}\left(-\Theta^{-1} Q \Theta N \Omega + N\Theta^{-1} Q \Theta \Omega + 2\Theta^{-1} M \Omega\right) \\ &= \operatorname{tr}\left(\left(\left[N, \Theta^{-1} Q \Theta\right] + 2\Theta^{-1} M\right) \Omega\right). \end{aligned}$$

Theorem 2 *Θ is a stationary point (or a critical point) of η if and only if*

$$\operatorname{tr}\left(\left(\left[N, \Theta^{-1} Q \Theta\right] + 2\Theta^{-1} M\right) \Omega\right) = 0, \quad \forall \Omega \in \mathcal{L}. \tag{20}$$

Corollary 1 *If $M = 0$ and N, $Q \in \mathcal{J}$ or N, $Q \in \mathcal{L}$ then, Θ is a stationary point of η iff*

$$\left[N, \Theta^{-1} Q \Theta\right] = 0.$$

Proof: We had already prove in lemma (7), that if $Q \in \mathcal{J}$ $(\mathcal{L})$ then $\Theta^{-1} Q \Theta \in \mathcal{J}$ $(\mathcal{L})$. But, according to (8), $[\mathcal{L}, \mathcal{L}] \subset \mathcal{L}$ and $[\mathcal{J}, \mathcal{J}] \subset \mathcal{L}$.

So, in cases Q, $N \in \mathcal{J}$ or Q, $N \in \mathcal{L}$,

$$\left[N, \Theta^{-1} Q \Theta\right] \in \mathcal{L},$$

and

$$\operatorname{tr}\left(\left[N, \Theta^{-1} Q \Theta\right] \Omega\right) = 0, \quad \forall \Omega \in \mathcal{L} \iff \left[N, \Theta^{-1} Q \Theta\right] = 0.$$

■

Remark 5 *Note that when $M = 0$, if $N \in \mathcal{L}$ and $Q \in \mathcal{J}$, or, $N \in \mathcal{J}$ and $Q \in \mathcal{L}$, every $\Theta \in G$ is a stationary point of η, since in these cases $T_\Theta \eta \equiv 0$.*

If we restrict ourselves to the Lie group G referred on the previous section, for the particular case when $P^2 = \pm I$, we note that G is a compact Lie group, with respect to the Riemannian metric

$$\ll \Theta\Omega_1, \Theta\Omega_2 \gg = \langle \Omega_1, \Omega_2 \rangle_P = \mathrm{tr}\left(\Omega_1^P \Omega_2\right), \quad \Omega_1,\ \Omega_2 \in \mathcal{L}. \tag{21}$$

From the compactness of the Lie group G, it makes sense to consider the following minimization problem:

$$(\mathcal{P}_1) \qquad \min_{\Theta \in G} \ \eta(\Theta).$$

The Riemannian metric (21) enables us to give another characterization of the stationary points of η.

Theorem 3 *If $P^2 = \pm I$, $\Theta \in G$ is a stationary point of η if and only if*

$$\left[N, \Theta^{-1} Q \Theta\right] + 2\Theta^{-1} M \in \mathcal{J}. \tag{22}$$

Proof: From lemmas 4, 6 and since $\Theta \in G$, condition (20) may be rewritten as

$$\mathrm{tr}\left(\left(\left[\Theta^{-1} Q^P \Theta, N^P\right] + 2M^P \Theta\right)^P \Omega\right) = 0, \quad \forall \Omega \in \mathcal{L}. \tag{23}$$

The definition of the Riemannian metric (21), the direct decomposition

$$\mathcal{G}l(n, \mathbb{R}) = \mathcal{L} \oplus \mathcal{J},$$

together with condition (23), implies that Θ is a stationary point of η if and only if

$$\left[\Theta^{-1} Q^P \Theta, N^P\right] + 2M^P \Theta \in \mathcal{J},$$

or, equivalently, due to lemma 5, if and only if

$$\left(\left[\Theta^{-1} Q^P \Theta, N^P\right] + 2M^P \Theta\right)^P = \left[N, \Theta^{-1} Q \Theta\right] + 2\Theta^{-1} M \in \mathcal{J}.$$

■

Corollary 2 *If $P^2 = \pm I$, $\left[N, \Theta^{-1} Q \Theta\right] = 0$ (in particular if $N = 0$ or $Q = 0$), Θ is a stationary point of η iff $\Theta^{-1} M \in \mathcal{J}$.*

Proof: Follows immediately from theorem 3. ■

Let us analyze separately some particular cases of the problem $(\mathcal{P}_1)$.

3.1 $P^2 = \pm I$, $N = 0$ (or $Q = 0$) and M nonsingular

In this case the generalized Brockett function takes the simplified form

$$\eta(\Theta) = -2\,\mathrm{tr}\left(M\Theta^{-1}\right). \tag{24}$$

According to corollary 2, if Θ is a stationary point of η, then

$$\begin{aligned}
\Theta^{-1}M \in \mathcal{J} \quad &\iff \quad \left(\Theta^{-1}M\right)^{\top} P = P\Theta^{-1}M \\
&\iff \quad M^{\top}\left(\Theta^{-1}\right)^{\top} P = \Theta^{\top} P M \\
&\iff \quad M^{\top} P\Theta = \Theta^{\top} P M. \qquad (25)
\end{aligned}$$

Note that, if $P^2 = I$,

$$(25) \iff M^{\top} P\Theta \text{ is symmetric,}$$

and if $P^2 = -I$,

$$(25) \iff M^{\top} P\Theta \text{ is skew} - \text{symmetric.}$$

The next theorem gives us an explicit form for the stationary points of the function η given by (24).

Theorem 4 *If $P^2 = \pm I$, the critical points of $\eta(\Theta) = -2\,\mathrm{tr}\left(M\Theta^{-1}\right)$ are given explicitly by*

$$\Theta^{-1}M = \sqrt{M^{P}M}, \tag{26}$$

where $\sqrt{M^{P}M}$ is any square root of $M^{P}M$ which belongs to $\mathcal{J}$.

Proof: Attending to (25) and the properties of the elements of G, we have

$$\begin{aligned}
M^{\top} P\Theta = \Theta^{\top} P M \quad &\iff \quad M^{\top} P\Theta = P\Theta^{-1}M \\
&\iff \quad P^{\top} M^{\top} P = \Theta^{-1} M \Theta^{-1}.
\end{aligned}$$

Since $P^{\top} M^{\top} P$ is nonsingular, $\Theta^{-1} \in G$ and $\Theta^{-1}M \in \mathcal{J}$, $P^{\top} M^{\top} P = \Theta^{-1} M \Theta^{-1}$ is the generalized polar decomposition of $P^{\top} M^{\top} P$ (see [4]). So,

$$\Theta^{-1}M = \sqrt{P^{\top} M^{\top} P \left(P^{\top} M^{\top} P\right)^{P}} = \sqrt{P^{\top} M^{\top} P M} = \sqrt{M^{P}M}.$$

■

It also happens that when M is invertible the matrix $M^{P}M$ has a unique square root whose spectrum lies in the open right half complex pane (see [4]). This, together with theorem 4 are now enough to characterize the minimum.

Theorem 5 *The function $\eta(\Theta) = -2\,\mathrm{tr}\left(M\Theta^{-1}\right)$, with M invertible, is minimized when*

$$\Theta^{-1}M = \left(M^P M\right)^{\frac{1}{2}}, \tag{27}$$

where $\left(M^P M\right)^{\frac{1}{2}}$ is the unique square root of $M^P M$ that belongs to $\mathcal{J}$ and has all the eigenvalues with positive real part.

By using the Riemannian metric defined on G by (21), we can pass from the linear term of the Taylor series expansion of a smooth function defined on G, to a vector field on G named the gradient vector field.

Definition 1 (Helmke and Moore [7])
Given a smooth function $f\colon G \longrightarrow \mathbb{R}$, defined on the Lie group G, the gradient vector field of f relative to the Riemannian metric $\ll \cdot,\cdot \gg$ on G, is the unique vector field of G, $\nabla f\colon G \longrightarrow TG$, which satisfies

$$T_\Theta f(\Theta\Omega) = \ll \nabla f(\Theta), \Theta\Omega \gg, \quad \forall \Omega \in \mathcal{L},$$

where $T_\Theta f$ stands for the tangent map of f at Θ.

We have already seen at the beginning of section 3 that in this case

$$T_\Theta \eta(\Theta\Omega) = 2\,\mathrm{tr}\left(\Theta^{-1}M\Omega\right).$$

According to definition 1, the gradient vector field of η, $\nabla\eta$, relative to the Riemannian metric (21), is the unique vector field of G satisfying

$$\ll \nabla\eta(\Theta), \Theta\Omega \gg = 2\,\mathrm{tr}\left(\Theta^{-1}M\Omega\right), \ \ \forall \Omega \in \mathcal{L}. \tag{28}$$

Now, since $\nabla\eta(\Theta) \in T_\Theta G$, we can write

$$\nabla\eta(\Theta) = \Theta X,$$

with $X \in \mathcal{L}$.

So, condition (28) is equivalent to

$$\mathrm{tr}\left(X^P\Omega\right) = 2\,\mathrm{tr}\left(\Theta^{-1}M\Omega\right), \ \ \forall \Omega \in \mathcal{L}, \tag{29}$$

and thus,

$$\langle 2M^P\Theta - X, \Omega\rangle_P = 0, \ \ \forall \Omega \in \mathcal{L} \iff 2M^P\Theta - X \in \mathcal{J}. \tag{30}$$

But,

$$2M^P\Theta - X \in \mathcal{J} \quad \iff \quad \left(2M^P\Theta - X\right)^\top P = P\left(2M^P\Theta - X\right)$$

$$\begin{aligned}
&\iff \quad 2\Theta^{\top}\left(M^{P}\right)^{\top}P - X^{\top}P = 2PM^{P}\Theta - PX \\
&\iff \quad 2\Theta^{\top}\left(P^{\top}M^{\top}P\right)^{\top}P + PX = 2PM^{P}\Theta - PX \\
&\iff \quad PX = PM^{P}\Theta - \Theta^{\top}P^{\top}MPP \\
&\iff \quad X = M^{P}\Theta - P^{\top}P^{\top}\Theta^{-1}MPP \\
&\iff \quad X = M^{P}\Theta - \Theta^{-1}M, \quad \left(P^{2} = \pm I\right)
\end{aligned}$$

and then we get

$$\nabla\eta(\Theta) = \Theta\left(M^{P}\Theta - \Theta^{-1}M\right).$$

Thus, the gradient flow equation associated to η, relatively to the Riemannian metric (21) is given by

$$\dot{\Theta} = \Theta M^{P}\Theta - M. \tag{31}$$

Note that the equilibrium points of the previous dynamical system are exactly the stationary points of the function η given by (24).

One of the questions that arises naturally at this moment is concerned about the integrability of the nonlinear ordinary differential equation (31). We are going to prove that for the particular case when $M \in \mathcal{J}(\mathcal{L})$, that is when $M^{P} = M$ ($M^{P} = -M$), the gradient flow equation (31) is completely integrable and we give here its explicit solution for every initial condition $\Theta(0) = \Theta_0 \in G$. Finally, we present an explicit solution of a linearization of (31), for an arbitrary nonsingular matrix M.

Definition 2 *The Cayley transform of a matrix Θ, not having -1 as an eigenvalue, is given by*

$$Y = (I - \Theta)(I + \Theta)^{-1}. \tag{32}$$

Theorem 6 *The Cayley transform (32) defines a bijection between elements of G and elements of $\mathcal{L}$ (not having -1 as an eigenvalue).*

Proof: Since $(I - \Theta)(I + \Theta) = (I + \Theta)(I - \Theta)$, we can also write

$$Y = (I + \Theta)^{-1}(I - \Theta). \tag{33}$$

Let us begin to prove that the spectrum of Y doesn't contain -1.

If there was a nonzero vector v such that $Yv = -v$, attending to (33),

$$Yv = -v \iff (I + \Theta)^{-1}(I - \Theta)v = -v \iff v - \Theta v = -v - \Theta v \iff v = 0,$$

which is a contradiction.

Then, the matrix $I+Y$ is invertible and,

$$\begin{aligned}Y=(I-\Theta)(I+\Theta)^{-1} \quad &\Longleftrightarrow \quad Y(I+\Theta)=I-\Theta\\ &\Longleftrightarrow \quad (I+Y)\Theta=I-Y\\ &\Longleftrightarrow \quad \Theta=(I+Y)^{-1}(I-Y).\end{aligned} \tag{34}$$

It remains to prove that $Y\in\mathcal{L}$ iff $\Theta\in G$.

$$\begin{aligned}Y\in\mathcal{L} \quad &\Longleftrightarrow \quad Y^{\top}P=-PY\\ &\Longleftrightarrow \quad \left((I+\Theta)^{-1}\right)^{\top}(I-\Theta)^{\top}P=-P(I-\Theta)(I+\Theta)^{-1}\\ &\Longleftrightarrow \quad (I-\Theta^{\top})P(I+\Theta)=-(I+\Theta^{\top})P(I-\Theta)\\ &\Longleftrightarrow \quad P-\Theta^{\top}P+P\Theta-\Theta^{\top}P\Theta=-P-\Theta^{\top}P+P\Theta+\Theta^{\top}P\Theta\\ &\Longleftrightarrow \quad \Theta^{\top}P\Theta=P\\ &\Longleftrightarrow \quad \Theta\in G.\end{aligned}$$

■

Lemma 10 (Veselov and Dynnikov [11])
If $M\in\mathcal{J}$ ($M^{P}=M$), the Cayley transform (32) linearizes the flow (31):

$$\dot{Y}=\{Y,M\}=YM+MY. \tag{35}$$

Proof: From (33), we can also write (34) as

$$\Theta=(I-Y)(I+Y)^{-1}. \tag{36}$$

Furthermore,

$$\begin{aligned}I+\Theta \quad &= \quad I+(I+Y)^{-1}(I-Y)\\ &= \quad (I+Y)^{-1}(I+Y+I-Y)\\ &= \quad 2(I+Y)^{-1},\end{aligned}$$

and thus

$$(I+\Theta)^{-1}=\frac{1}{2}(I+Y). \tag{37}$$

Differentiating equation (32), we derive

$$\begin{aligned}\dot{Y} \quad &= \quad -\dot{\Theta}(I+\Theta)^{-1}-(I-\Theta)(I+\Theta)^{-1}\dot{\Theta}(I+\Theta)^{-1}\\ &= \quad -\left(I+(I-\Theta)(I+\Theta)^{-1}\right)\dot{\Theta}(I+\Theta)^{-1}\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{2}(I+Y)(\Theta M\Theta - M)(I+Y) \quad \text{(by (32), (31) and (37))} \\
&= -\frac{1}{2}(I+Y)\Big((I+Y)^{-1}(I-Y)M(I-Y)(I+Y)^{-1} - M\Big)(I+Y) \quad \text{(by (34) and (36))} \\
&= -\frac{1}{2}((I-Y)M(I-Y) - (I+Y)M(I+Y)) \\
&= YM + MY.
\end{aligned}
$$

■

Remark 6 *Note that, since $\Theta \in G$, from theorem 6 we conclude that $Y \in \mathcal{L}$ and thus, equation (35) evolves on the Lie algebra $\mathcal{L}$ if and only if $M \in \mathcal{J}$ (see condition (7)).*

Theorem 7 (Veselov and Dynnikov [11])
If $M \in \mathcal{J}$, equation (31) can be solved explicitly for every initial condition $\Theta(0) = \Theta_0 \in G$. The solution has the form

$$
\Theta(t) = (-\sinh(tM) + \cosh(tM)\Theta_0)(\cosh(tM) - \sinh(tM)\Theta_0)^{-1}.
$$

Proof: The explicit solution of the linear differential equation (35) for the initial point $Y_0 = (I - \Theta_0)(I + \Theta_0)^{-1}$ is given by

$$
Y(t) = \exp(tM) Y_0 \exp(tM), \tag{38}
$$

where $\exp(\cdot)$ denotes the exponential map on the Lie algebra $\mathcal{G}l(n, \mathbb{R})$.

Since

$$
\Theta(t) = (I - Y(t))(I + Y(t))^{-1},
$$

by making use of (38), we derive

$$
\begin{aligned}
I - Y(t) &= \exp(-tM)\exp(tM) - \exp(tM)(I - \Theta_0)(I + \Theta_0)^{-1}\exp(tM) \\
&= (\exp(-tM)(I + \Theta_0) - \exp(tM)(I - \Theta_0))(I + \Theta_0)^{-1}\exp(tM) \\
&= (\exp(-tM) + \exp(-tM)\Theta_0 - \exp(tM) + \exp(tM)\Theta_0)(I + \Theta_0)^{-1}\exp(tM) \\
&= 2(-\sinh(tM) + \cosh(tM)\Theta_0)(I + \Theta_0)^{-1}\exp(tM),
\end{aligned}
$$

and analogously

$$
I + Y(t) = 2(\cosh(tM) - \sinh(tM)\Theta_0)(I + \Theta_0)^{-1}\exp(tM).
$$

Thus,

$$
\begin{aligned}
\Theta(t) &= 2(-\sinh(tM) + \cosh(tM)\Theta_0)(I + \Theta_0)^{-1}\exp(tM) \\
&\quad \frac{1}{2}\exp(-tM)(I + \Theta_0)(\cosh(tM) - \sinh(tM)\Theta_0)^{-1} \\
&= (-\sinh(tM) + \cosh(tM)\Theta_0)(\cosh(tM) - \sinh(tM)\Theta_0)^{-1}.
\end{aligned}
$$

■

In order to present the explicit solution of equation (31) for the case when $M \in \mathcal{L}$, we need to imbed this equation in the complex Lie group

$$G^{\star} = \{X \in GL(n,\mathbb{C}) \, : \, X^{\star}PX = P\}\,, \tag{39}$$

where $X^{\star}$ denotes the transpose conjugate of X.

The corresponding Lie algebra and Jordan algebra are represented, respectively, by $\mathcal{L}^{\star}$ and $\mathcal{J}^{\star}$ and are given by

$$\mathcal{L}^{\star} = \{\Omega \in \mathcal{G}l(n,\mathbb{C}) \, : \, \Omega^{\star}P = -P\Omega\}\,, \tag{40}$$

$$\mathcal{J}^{\star} = \{A \in \mathcal{G}l(n,\mathbb{C}) \, : \, A^{\star}P = PA\}\,. \tag{41}$$

Lemma 11 *If $\Theta \in G$ then $i\Theta \in G^{\star}$.*

Proof: Since $\Theta \in G$, Θ has real entries and satisfies $\Theta^{\top}P\Theta = P$.

So,

$$(i\Theta)^{\star}P(i\Theta) = -i\Theta^{\top}Pi\Theta = \Theta^{\top}P\Theta = P.$$

■

Lemma 12 *If $A \in \mathcal{L}$ ($\mathcal{J}$) then $iA \in \mathcal{J}^{\star}$ ($L^{\star}$).*

Proof: Let us suppose that $A \in \mathcal{L}$. The other case is proved in an analogue way.

Then, $A^{\top}P = -PA$, and

$$(iA)^{\star}\, P = -iA^{\top}P = P\,(iA)\,,$$

which implies that $iA \in \mathcal{J}^{\star}$. ■

We now introduce a very useful change of variable which recast equation (31) as one which evolves in $G^{\star}$.

Let

$$X = i\Theta. \tag{42}$$

Differentiating both sides of equation (42) in order to t, and making use of (31), we derive

$$\begin{aligned} \dot{X} = i\dot{\Theta} \quad &= \quad i\,(-\Theta M\Theta - M) \\ &= \quad i\,(XMX - M) \\ &= \quad X(iM)X - (iM). \end{aligned} \tag{43}$$

Theorem 8 *If $M \in \mathcal{L}$ ($M^{P} = -M$), the explicit solution of equation (31) for the initial condition $\Theta(0) = \Theta_0$ has the form*

$$\Theta(t) = (-\sin(tM) + \cos(tM)\Theta_0)\,(\cos(tM) + \sin(tM)\Theta_0)^{-1}\,. \tag{44}$$

Proof: According to theorem 7, the explicit solution of the differential equation (43) for the initial condition $i\Theta_0 = X_0 \in G^\star$ is

$$X(t) = \left(-\sinh(itM) + \cosh(itM)X_0\right)\left(\cosh(itM) - \sinh(itM)X_0\right)^{-1}$$

Hence, according to (42), we obtain

$$\begin{aligned}\Theta(t) &= -i\left(-\sinh(itM) + i\cosh(itM)\Theta_0\right)\left(\cosh(itM) - i\sinh(itM)\Theta_0\right)^{-1} \\ &= \left(-\sin(tM) + \cos(tM)\Theta_0\right)\left(\cos(tM) + \sin(tM)\Theta_0\right)^{-1}.\end{aligned}$$

■

By setting $\Theta = XY^{-1}$, equation (31) is equivalent to the system of linear differential equations

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \end{bmatrix} = \begin{bmatrix} 0 & -M \\ -M^P & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix}. \tag{45}$$

For the initial condition $\begin{bmatrix} X_0 & Y_0 \end{bmatrix}^\top$, the solution of (3.1) is

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \exp\left(tA\right) \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix},$$

where

$$A = -\begin{bmatrix} 0 & M \\ M^P & 0 \end{bmatrix},$$

and $\exp(tA) = I + \sum_{n=1}^{+\infty} \frac{A^n}{n!} t^n$.

By computing the successive powers of A, we get

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cosh\left(t\sqrt{MM^P}\right) & -\left(\sqrt{MM^P}\right)^{-1} \sinh\left(t\sqrt{MM^P}\right) M \\ -M^P \sinh\left(t\sqrt{MM^P}\right)\left(\sqrt{MM^P}\right)^{-1} & \cosh\left(t\sqrt{M^P M}\right) \end{bmatrix} \begin{bmatrix} X_0 \\ Y_0 \end{bmatrix}.$$

3.2 $P^2 = \pm I$, $M = 0$ and $N,\ Q \in \mathcal{J}$ (or $N,\ Q \in \mathcal{L}$)

In this case, the generalized Brockett function takes the form

$$\eta(\Theta) = \operatorname{tr}\left(\Theta^{-1} Q \Theta N\right). \tag{46}$$

Attending to (23), $\Theta \in G$ is a stationary point of (46) iff

$$\left[\Theta^{-1} Q^P \Theta, N^P\right] = 0,$$

which is equivalent to

$$\left[N, \Theta^{-1} Q \Theta\right] = 0.$$

Lemma 13 *If $P^2 = \pm I$ and $N,\, Q \in \mathcal{J}$ (or $N,\, Q \in \mathcal{L}$), then*

$$\left[N, \Theta^{-1}Q\Theta\right] = 0 \iff N\Theta^{-1}Q\Theta \in \mathcal{J}.$$

Proof: Let us suppose that $N,\, Q \in \mathcal{J}$. (The proof is similar if $N,\, Q \in \mathcal{L}$.)

By definition of $\mathcal{J}$, if $A \in \mathcal{J}$ then

$$A = P^{\top}A^{\top}P.$$

So,

$$\begin{aligned}
\left[N, \Theta^{-1}Q\Theta\right] = 0 \quad &\iff \quad N\Theta^{-1}Q\Theta = \Theta^{-1}Q\Theta N \iff PN^{\top}P\Theta^{-1}Q\Theta = \Theta^{-1}PQ^{\top}P\Theta N \\
&\iff \quad PN^{\top}\Theta^{\top}PQ\Theta = P\Theta^{T}Q^{\top}P\Theta N \iff N^{\top}\Theta^{\top}Q^{\top}P\Theta = \Theta^{T}PQ\Theta N \\
&\iff \quad N^{\top}\Theta^{\top}Q^{\top}\left(\Theta^{-1}\right)^{\top}P = P\Theta^{-1}Q\Theta N \iff \left(\Theta^{-1}Q\Theta N\right)^{\top}P = P\Theta^{-1}Q\Theta N \\
&\iff \quad \Theta^{-1}Q\Theta N \in \mathcal{J}.
\end{aligned}$$

■

Theorem 9 *The gradient flow equation of $\eta(\Theta) = \operatorname{tr}\left(\Theta^{-1}Q\Theta N\right)$, relatively to the Riemannian metric (21), is given, by*

$$\dot{\Theta} = \Theta\left[N, \Theta^{-1}Q\Theta\right]^{P}. \tag{47}$$

Proof: In fact, the tangent map of $\eta(\Theta) = \operatorname{tr}\left(\Theta^{-1}Q\Theta N\right)$ at Θ, is, according to theorem 2, given by

$$\begin{aligned}
T_{\Theta}\eta(\Theta\Omega) \quad &= \quad \operatorname{tr}\left(\left[N, \Theta^{-1}Q\Theta\right]\Omega\right) \\
&= \quad \operatorname{tr}\left(\left[\Theta^{-1}Q^{P}\Theta, N^{P}\right]^{P}\Omega\right), \quad \forall \Omega \in \mathcal{L}.
\end{aligned}$$

On the other hand, the gradient vector field of η, $\nabla\eta$, relatively to the Riemannian metric (21), is the unique vector field of G satisfying

$$\ll \nabla\eta(\Theta), \Theta\Omega \gg = \operatorname{tr}\left(\left[\Theta^{-1}Q^{P}\Theta, N^{P}\right]^{P}\Omega\right), \ \forall \Omega \in \mathcal{L}.$$

Now, since $[\mathcal{J}, \mathcal{J}] \subset \mathcal{L}$ and $[\mathcal{L}, \mathcal{L}] \subset \mathcal{L}$,

$$\left[N, \Theta^{-1}Q\Theta\right] \in \mathcal{L},$$

then

$$\begin{aligned}
\nabla\eta(\Theta) \quad &= \quad \Theta\left[\Theta^{-1}Q^{P}\Theta, N^{P}\right] \\
&= \quad \Theta\left[N, \Theta^{-1}Q\Theta\right]^{P}
\end{aligned}$$

The gradient flow equation is then given by

$$\dot{\Theta} = \nabla \eta(\Theta),$$

and the result follows. ■

The dynamical system (47) evolves on the Lie group of $P-$orthogonal matrices and the equilibrium points are exactly the stationary points of the function η.

We will see that with a useful change of variables, equation (47), which evolves on the Lie group G, can be rewrite as a equation, actually a double bracket equation, evolving on the space of $P-$symmetric ($P-$skew-symmetric) matrices.

Let

$$H = \Theta^{-1} Q^P \Theta. \tag{48}$$

If we differentiate both sides of equation (48) with respect to t we get

$$\begin{aligned}
\dot{H} &= (\dot{\Theta^{-1}}) Q^P \Theta + \Theta^{-1} Q^P \dot{\Theta} \\
&= -\Theta^{-1} \dot{\Theta} \Theta^{-1} Q^P \Theta + \Theta^{-1} Q^P \dot{\Theta} \\
&= -\Theta^{-1} \Theta \left[N, \Theta^{-1} Q \Theta\right]^P \Theta^{-1} Q^P \Theta + \Theta^{-1} Q^P \Theta \left[N, \Theta^{-1} Q \Theta\right]^P \\
&= \left[\Theta^{-1} Q^P \Theta, N^P\right] \Theta^{-1} Q^P \Theta + \Theta^{-1} Q^P \Theta \left[\Theta^{-1} Q^P \Theta, N^P\right] \\
&= -\left[H, N^P\right] H + H \left[H, N^P\right] \\
&= \left[H, \left[H, N^P\right]\right].
\end{aligned}$$

So, as expected, we reduce equation (47) to the isospectral flow given by the double Lie bracket equation

$$\dot{H} = \left[H, \left[H, N^P\right]\right]. \tag{49}$$

Remark 7 *From lemmas 5 and 7, we note that $H \in \mathcal{J}(\mathcal{L}) \iff Q \in \mathcal{J}(\mathcal{L})$, and thus, according to (8), if N, $Q \in \mathcal{J}(\mathcal{L})$, equation (49) evolves on the space of $P-$symmetric ($P-$skew-symmetric) matrices.*

Interesting standard computational problems can be solved using this double bracket equation and are given in Brockett [3], for the particular case when $P = I$ and N and Q are symmetric matrices with unrepeated eigenvalues. Equation (49) can be used, between other things, to sort lists of real numbers, to diagonalize symmetric matrices and to solve some linear programming problems.

Bloch [1] also relate this gradient flow equation with completely integrable Hamiltonian systems, like the Toda flow on tridiagonal matrices, when studied from a "Lax pair" point of view, and the total least squares problem (the problem of fitting planes to a given data set of points).

Furthermore, working with various discretizations of equation (49), it provides a way to obtain even more stable and perhaps faster algorithms that converge to the solution of the desired problem.

To finish this section, we want to notify that we could compute the gradient flow equation for the generalized Brockett function without any particularization of matrices N, Q or M.

Following the spirit of Brockett's work [2], we present a result that enables us to write the desired gradient flow equation, relatively to the Riemannian metric (21).

Theorem 10 *Consider G (or any Lie subgroup of G) as a Riemannian manifold endowed with the Riemannian metric (21). The gradient flow on G corresponding to the generalized Brockett function*

$$\eta(\Theta) = \operatorname{tr}\left(\Theta^{-1}Q\Theta N - 2M\Theta^{-1}\right)$$

may be expressible as

$$\dot{\Theta} = \Theta \sum_{i=1}^{\dim \mathcal{L}} \operatorname{tr}\left(\left(\left[N, \Theta^{-1}Q\Theta\right] + 2\Theta^{-1}M\right)\Omega_i\right)\Omega_i, \tag{50}$$

where $\{\Omega_1, \ldots, \Omega_{\dim \mathcal{L}}\}$ is any orthonormal basis for $\mathcal{L}$.

Proof: Attending to (23) and the definition of the gradient flow equation, we must have

$$\operatorname{tr}\left(\left(\left[\Theta^{-1}Q^P\Theta, N^P\right] + 2M^P\Theta\right)^P \Omega\right) = \ll \nabla\eta(\Theta), \Theta\Omega \gg, \ \ \forall \Omega \in \mathcal{L},$$

or equivalently,

$$\langle \left[\Theta^{-1}Q^P\Theta, N^P\right] + 2M^P\Theta, \Omega\rangle_P = \langle X, \Omega\rangle_P, \ \ \forall \Omega \in \mathcal{L}, \tag{51}$$

where $\nabla\eta(\Theta) = \Theta X$, and $X \in \mathcal{L}$.

Since $\{\Omega_1, \ldots, \Omega_{\dim \mathcal{L}}\}$ is an orthonormal basis for $\mathcal{L}$, we can write

$$X = \sum_{i=1}^{\dim \mathcal{L}} \alpha_i \Omega_i, \quad \alpha_i \in \mathbb{R},$$

and equation (51) may be rewritten as

$$\langle \left[\Theta^{-1}Q^P\Theta, N^P\right] + 2M^P\Theta, \Omega_j\rangle_P = \sum_{i=1}^{\dim \mathcal{L}} \alpha_i \langle \Omega_i, \Omega_j\rangle_P, \ \ j = 1, \ldots, \dim \mathcal{L}. \tag{52}$$

Then,

$$\alpha_i = \langle \left[\Theta^{-1}Q^P\Theta, N^P\right] + 2M^P\Theta, \Omega_i\rangle_P,$$

and the result follows. ■

The purpose of exhibit the gradient flow of the generalized Brockett function for the particular cases of matrices Q, N and M studied on the previous two subsections was to present simplified forms of equation (50) that have a great numerical interest.

They could be used to solve certain computational problems and with the various discretizations of those continuous time equations we can obtain new algorithms that converge to the solution of the required problem.

4 Optimization matrix problems

In this section we present some optimization matrix problems that can be solved by using the generalized Brockett function presented in the previous section.

The nondegenerate symmetric bilinear form (17) defined in section 2 can be extended to the Lie algebra $\mathcal{G}l(n,\mathbb{R})$ as

$$\langle X,Y\rangle_P = \operatorname{tr}\left(X^P Y\right), \;\; X,\, Y \in \mathcal{G}l(n,\mathbb{R}). \tag{53}$$

According to remark 3, the bilinear form (53) carries out a pseudo-Riemannian structure to the manifold of nonsingular matrices $GL(n,\mathbb{R})$ that can be used to find the $P-$orthogonal matrix that is closest to a given nonsingular matrix, in a certain sense.

Let M be a nonsingular n by n matrix, and consider the following optimization problem

$$(\mathcal{P}_2) \qquad \min_{\Theta\in G} \; \langle M-\Theta, M-\Theta\rangle_P \; .$$

Attending to the definition of the bilinear form (17), we have

$$\begin{aligned}
\langle M-\Theta, M-\Theta\rangle_P &= \operatorname{tr}\left(P^\top\left(M^\top-\Theta^\top\right)P(M-\Theta)\right)\\
&= \operatorname{tr}\left(P^\top M^\top PM - P^\top\Theta^\top PM - P^\top M^\top P\Theta + P^\top\Theta^\top P\Theta\right)\\
&= \langle M,M\rangle_P - \operatorname{tr}\left(\Theta^{-1}P^\top PM + \Theta^\top P^\top MP\right) + n\\
&= \langle M,M\rangle_P - 2\operatorname{tr}\left(\Theta^{-1}M\right) + n.
\end{aligned}$$

Thus,

$$(\mathcal{P}_2) \iff \min_{\Theta\in G} \; -2\operatorname{tr}\left(\Theta^{-1}M\right).$$

Note that the right hand side is nothing else than the generalized Brockett function for the particular case when $N=0$ or $Q=0$, and was already studied in subsection 3.1.

Attending to theorem 7, the solution of this optimization problem is given by

$$\Theta^{-1}M = \left(M^P M\right)^{\frac{1}{2}},$$

where $\left(M^P M\right)^{\frac{1}{2}}$ denotes the unique square root of $M^P M$ that belongs to $\mathcal{J}$ and has all the eigenvalues with positive real parts.

If we consider the generalized polar decomposition of the matrix M,

$$M = R\Lambda,$$

where $R \in G$ and $\Lambda \in \mathcal{J}$ $(\sigma(\Lambda) \subset \mathbb{C}^+)$, the solution of $(\mathcal{P}_2)$ is

$$\begin{aligned}
\Theta &= R\Lambda \left(P^\top \Lambda^\top R^\top P R \Lambda\right)^{-\frac{1}{2}} \\
&= R\Lambda \left(P^\top \Lambda^\top P \Lambda\right)^{-\frac{1}{2}} \\
&= R\Lambda \left(P^\top P \Lambda^2\right)^{-\frac{1}{2}} \\
&= R\Lambda\Lambda^{-1} \\
&= R,
\end{aligned}$$

which generalizes the well known result that the orthogonal matrix that is nearest to a given nonsingular matrix is the orthogonal part of its polar decomposition.

Another interesting optimization problem is presented in what follows.

Let us consider the smooth action of the Lie group G on $\mathcal{G}l(n, \mathbb{R})$:

$$\begin{array}{rccc}
\phi\colon & G \times \mathcal{G}l(n,\mathbb{R}) & \longrightarrow & \mathcal{G}l(n,\mathbb{R}) \\
& (\Theta, A) & \longmapsto & \Theta A \Theta^{-1}
\end{array}.$$

In fact,

$$\phi(I, A) = A,$$

and

$$\phi\left(\Theta_1, \phi(\Theta_2, A)\right) = \phi\left(\Theta_1, \Theta_2 A \Theta_2^{-1}\right) = \Theta_1 \Theta_2 A \left(\Theta_1 \Theta_2\right)^{-1} = \phi(\Theta_1 \Theta_2, A), \quad \Theta_1,\ \Theta_2 \in G.$$

Given a $P-$symmetric $n \times n$ matrix Q, the set of the matrices that are $P-$orthogonally similar to Q,

$$\mathcal{M}(Q) = \left\{\Theta^{-1} Q \Theta \ :\ \Theta \in G\right\}, \tag{54}$$

is an orbit under the group action ϕ and therefore is a smooth manifold of $\mathcal{G}l(n, \mathbb{R})$.

If N is an arbitrary $P-$symmetric matrix , we can consider the following matrix function

$$\begin{array}{rccl}
f_N\colon & \mathcal{M}(Q) & \longrightarrow & \mathbb{R} \\
& X & \longmapsto & f_N(X) = \langle N - X, N - X\rangle_P
\end{array}.$$

Our objective is to find the $P-$symmetric matrix similar to Q, that is closest to the $P-$symmetric matrix N, relatively to (53).

We can formulate our problem like this:

$$(\mathcal{P}_3) \qquad\qquad \min_{X \in \mathcal{M}(Q)} \langle N - X, N - X\rangle_P \ .$$

Making use of (53), we can write

$$\langle N - X, N - X\rangle_P \ = \ \operatorname{tr}\left((N - X)^P (N - X)\right)$$

$$
\begin{aligned}
&= \operatorname{tr}\left(P^{\top}\left(N^{\top}-X^{\top}\right)P\left(N-X\right)\right) \\
&= \langle N,N\rangle_P - \operatorname{tr}\left(P^{\top}X^{\top}PN + P^{\top}N^{\top}PX\right) + \langle X,X\rangle_P \\
&= \langle N,N\rangle_P - \operatorname{tr}\left(P^{\top}PXN + P^{\top}PNX\right) + \langle X,X\rangle_P \\
&= \langle N,N\rangle_P - 2\operatorname{tr}\left(XN\right) + \langle X,X\rangle_P.
\end{aligned}
$$

Since $X \in \mathcal{M}(Q)$, it has the same spectrum as Q so, $\langle X,X\rangle_P$ is constant. Thus,

$$
(\mathcal{P}_3) \quad \Longleftrightarrow \quad \min_{X\in\mathcal{M}(Q)} \operatorname{tr}\left(-2XN\right) \quad \Longleftrightarrow \quad \min_{\Theta\in G} -\operatorname{tr}\left(\Theta^{-1}Q\Theta N\right),
$$

where $\operatorname{tr}\left(\Theta^{-1}Q\Theta N\right)$ is the generalized Brockett function with $M=0$ and was analyzed in subsection 3.2.

Remark 8 *A similar problem to $(\mathcal{P}_3)$ can be formulated if we consider both the matrices Q and N $P-$skew-symmetric. In that case, our problem becomes*

$$
(\mathcal{P}_4) \qquad \min_{\Theta\in G} \operatorname{tr}\left(\Theta^{-1}Q\Theta N\right).
$$

References

[1] Bloch, A. M., Steepest Descent, Linear Programming and Hamiltonian Flows, *Contemporary Mathematics*, 114: 77-88, (1990).

[2] Brockett, W. Roger, Least Squares Matching Problems, *Linear Algebra and its Applications*, 122/123/124: 761-777, (1989).

[3] Brockett, W. Roger, Dynamical Systems that Sort Lists, Diagonalize Matrices and Solve Linear Programming Problems, *Linear Algebra and its Applications*, 146: 79-91, (1991).

[4] Cardoso, J. R. and Silva Leite, F., Extending Results from Orthogonal to $P-$orthogonal Matrices. *To appear.*

[5] Chu, M., A List of Matrix Flowswith Applications, *Fields Institute Communications.*, 3 (1994) 87-97, AMS.

[6] Helgason, S., *Differential Geometry and Symmetric Spaces*, Academic Press, (1962).

[7] Helmke, U. and Moore, J., *Optimization and Dynamical Systems*, CCES, Springer-Verlag, London, (1994).

[8] Hüper, K., *Structure and Convergence of Jacobi-type Methods for Matrix Computations*, PhD Thesis, Techinal University of Munich, Germany, (1996).

[9] Munthe-Kaas, H., R. G. W. Quispel and A. Zanna, The Polar Decomposition on Lie Groups with Involutive Automorphisms, *Reports in Informatics*, 191, (2000).

[10] Silva Leite, F. and Crouch, P., Closed Forms for the Exponential Mapping on Matrix Lie Groups Based on Putzer's Method, *Journal of Mathematical Physics*, 40: 3561-3568, (1999).

[11] Veselov, A. P. and Dynnikov, I. A., Integrable Gradient Flows and Morse Theory, *St. Petersburg Math. J.*, 8 (No. 3): 429-446, (1997).

[12] Wong, W. S., Gradient Flows for Local Minima of Combinatorial Optimization problems, *Fields Institute Communications.*, 3 (1994) 145-155, AMS.